International Journal of Geology, Earth & Environmental Sciences ISSN: 2277-2081
An Open Access, Online International Journal Available at http://www.cibtech.org/jgee.htm
2025 Vol. 15, pp. 222-230/Diyorbek et al.
Research Article

ENGINEERING-SEISMOLOGICAL PLANNING OF OBSERVATION POINTS FOR VS30 MAPPING VIA POISSON-DISK SAMPLING UNDER ACCESS CONSTRAINTS

Diyorbek Yusupov [a,b], Ilhom Alimuhamedov [a], Suratbek Khalbayev [a,b]
[a] *G.O. Mavlonov Institute of Seismology, Academy of Sciences of the Republic of Uzbekistan*
[b] *Branch of the Gubkin Russian State University of Oil and Gas in Tashkent, Republic of Uzbekistan*

ABSTRACT
We present an engineering-seismological workflow for planning observation points in Vs30 mapping under access and safety constraints. The design is based on Poisson-disk sampling (PDS), which enforces a minimum inter-point distance to avoid clustering and introduces a practical service radius rrr that allows safe local relocation around restricted zones. On a controlled 100×100 synthetic testbed with four litho-zones, we compare random, regular grid, and PDS layouts at sparse designs (N = 25, 15, 7). Interpolation uses ordinary kriging and machine-learning models (SVR, XGBoost, k-NN) trained with Residual (RF) Bootstrap augmentation. PDS delivers more uniform coverage, fewer interpolation artifacts, and consistently better accuracy/robustness (MAE/RMSE/correlation/R^2), with the strongest gains at $N \leq 15$. The results support PDS as a default survey-planning tool for engineering microzonation and site-class assignment, improving field logistics while preserving uniform coverage around constrained infrastructure.

Keywords: *Vs30; Microzonation; Poisson-disk Sampling; ML-Interpolation; Kriging; Support Vector Regression; Access Constraints; Site Effects*

INTRODUCTION
The Vs30 parameter, which reflects the average shear wave velocity in the upper 30 m of the section, is a basic indicator in the engineering seismic assessment of territories and microzoning of particularly important facilities (e.g., nuclear power plants, hydraulic structures, bridges, and strategic industrial complexes). In such areas, the planning of geophysical surveys is strictly regulated by safety conditions, site accessibility, and the need to minimize work time, which imposes severe restrictions on the number and location of measurement points.

When there is a lack of field data for constructing VS30 maps, spatial interpolation methods (geostatistical Kriging) and machine learning approaches (Random Forest, SVR, XGBoost) are widely used. However, the effectiveness of these methods critically depends on the geometry of the points: classical schemes — random sampling or uniform grid — with a small number of measurements either create unnecessary clustering or leave “gaps” in the coverage of lithological bodies with sharp parameter gradients.

To eliminate these shortcomings, this work implements and investigates the Poisson Disk Sampling algorithm. The key advantage of the proposed approach is the functional ability to bypass “special” areas at specific sites due to the service radius of each point. This feature greatly simplifies the planning and logistics of field work, ensures stable interpolation, and reduces the risk of distortion when restoring the Vs30 parameter in conditions of limited access and data scarcity.

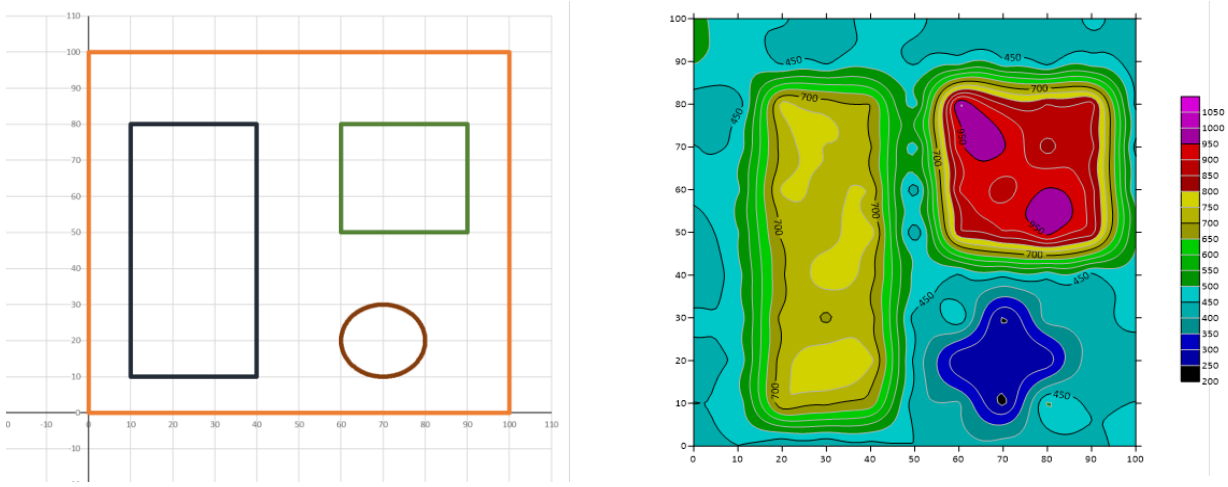
A practical study of the algorithm was conducted on a model map, where the Poisson disk sampling scheme was compared with classical approaches (random and regular placement) and evaluated using MAE, RMSE, MAPE, and other metrics for ML interpolation with RF bootstrap augmentation. The results demonstrate that the proposed algorithm provides stable and accurate restoration of even the Vs30 field in the presence of data scarcity, making it a promising tool for planning seismic surveys in the presence of special and particularly important objects.

MATERIALS AND METHODS

We compare three observation placement strategies: random (as a baseline), regular (as an engineering-intuitive approach), and Poisson-disk sampling, PDS (as providing quasi-regular coverage and logistical stability to restricted areas). The goal is to understand how the shape of the network at small N affects the stability of Vs30 maps and accuracy metrics. This focus is consistent with the practice of engineering seismology, where site conditions proxies (e.g., slope steepness and geomorphological classification) are widely used for preliminary estimates as indicators of Vs30 variations and soil amplification properties [1, 2].

2.1. Model map with geometrically regular lithological zones

A model geological map measuring 100×100 cells was used to analyze the effectiveness of various observation point placement schemes and subsequent interpolation of the Vs30 parameter. The spatial structure of the map was formed using four geometrically simple objects: a background area, a rectangle, a square, and a circle, each of which was considered as an independent lithological zone with its own velocity characteristics.



a. b.

Figure 1. Modeling of the Vs30 map by constructing simple geological figures. (a. – geometric distribution of geological zones; b. – distribution of Vs30 parameters according to the normal law, constructed using the Kriging method).

The Vs30 values within each zone were set according to a normal distribution with different medians and standard deviations reflecting typical values for common soil conditions. The statistical parameters were selected based on data for the Fergana Valley and the NEHRP classification, which ensured the realism and reproducibility of the model.

This approach allowed us to set a controlled structure with clearly defined boundaries between zones and was used as a benchmark for evaluating the quality of interpolation when varying the density and placement patterns of points. The resulting Vs30 field obtained by the Kriging method was further used to compare the accuracy of recovery when using different algorithms.

2.2. Method for distributing points on a map

2.2.1. Random distribution

As part of this strategy, a specified number of points (45, 35, 25, 15, and 7) were randomly distributed throughout the study area. The spatial distribution of points for each case is shown in Fig. 2.

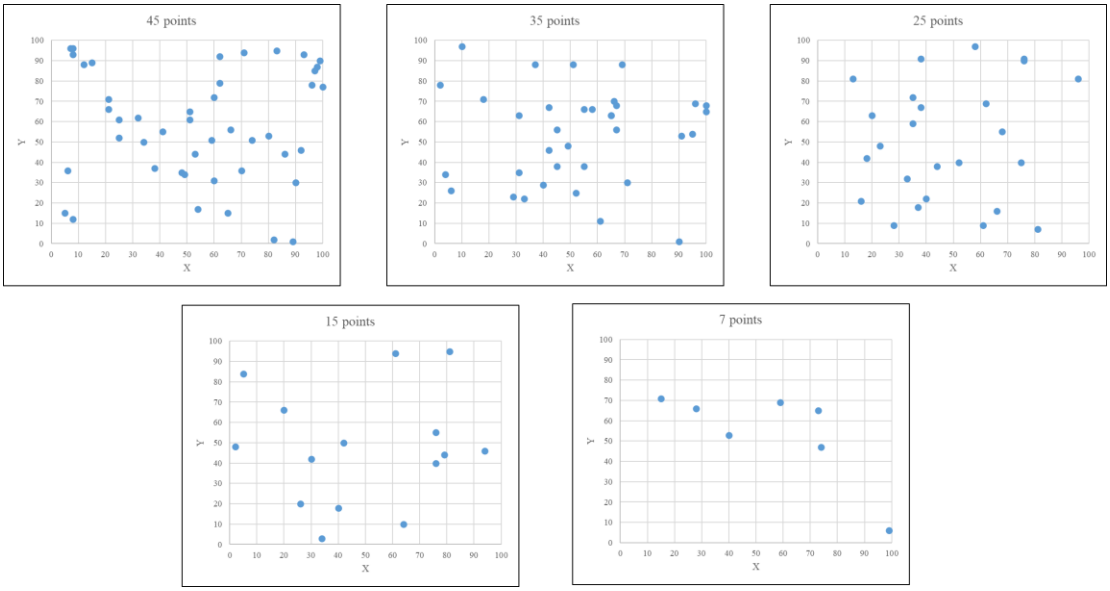


Figure 2. Spatial distribution of points on a map with random placement.

The configurations obtained served as the basis for interpolation using the conventional kriging method, the results of which are shown in Fig. 3.

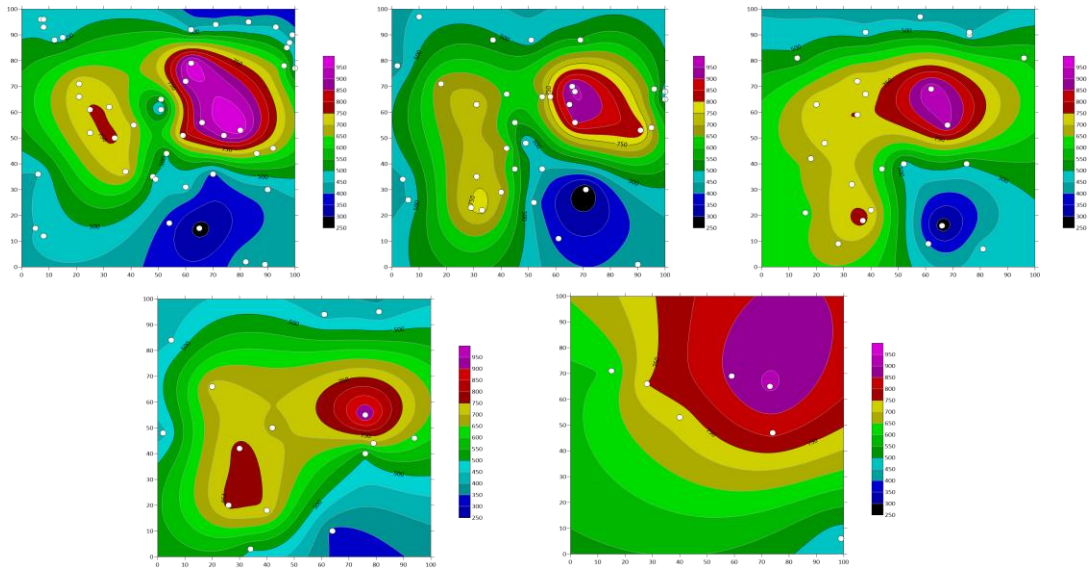


Figure 3. Results of interpolation using the k-nearest neighbors' method for random distribution of points (from 45 to 7 points) At high density (45 and 35 points), a smooth restoration of the spatial distribution is observed, with good coverage of both central and peripheral areas. However, as the number of points decreases (especially at 15 and 7), sharp gradients, artifacts, and underinterpolated areas begin to appear. This indicates a decrease in the stability and accuracy of spatial reproduction with a random but sparse sample distribution.

2.2.2. Regular distribution

To ensure uniform coverage of the territory with measurement points, an algorithm was used to divide the existing map into a series of cells of equal size with sides Δx and Δy, where $\Delta x=\Delta y$. This task boils down to the task of optimal discrete division of a two-dimensional map of fixed size into equal cells [3].

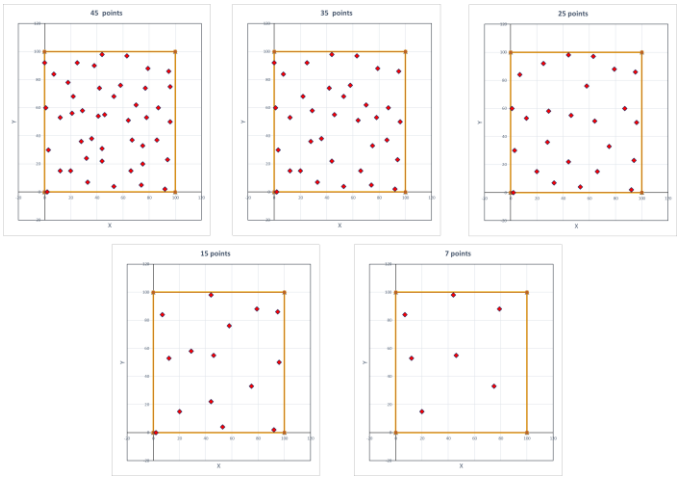


Figure 4. Algorithm for uniform coverage of observation points: 45, 35, 25, 15, 7 points for interpolation, and 10 points for validation

It should be noted that this method has direct practical significance: as a first approximation, it allows the formation of optimal point layout schemes for subsequent real-world planning of survey work, and also lays the foundation for automating the process of constructing observation point networks [3].
Thus, for each density of the main set of points, an additional fixed grid of selected points was used for independent verification of the interpolation quality.

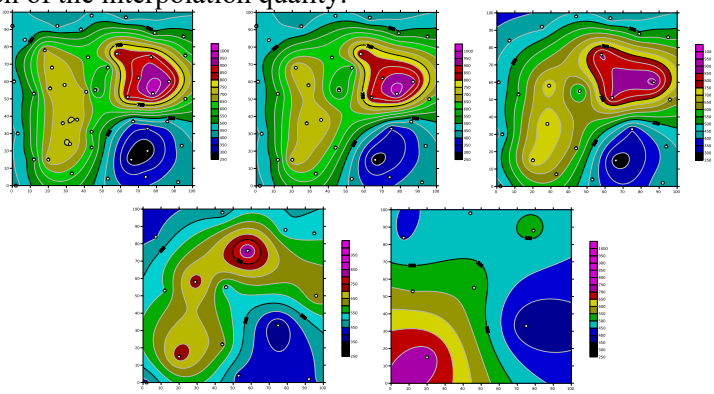


Figure 5. Results of kriking interpolation with regular distribution of points (from 45 to 7 points) When using a regular distribution of observation points (Fig. 5) with a high density of points (45 and 35), the regular grid provides uniform coverage of the territory, which allows accurate reproduction of the spatial distribution of Vs30 with smooth gradients and clear boundaries of lithological zones. However, when the number of points is reduced to 15 and 7, the quality of interpolation deteriorates: local distortions, sharp changes in values, and areas with insufficient detail appear, which indicates the limited effectiveness of the regular scheme in conditions of sparse sampling.

2.2.3. Poison Disk Sampling distribution

The Poisson disk sampling algorithm is based on the idea of placing points in such a way that the minimum distance between any two of them is maintained, as specified by a parameter, which ensures uniform coverage of the space without overlap or clustering. The space is represented as a two-dimensional square of fixed size, inside which an auxiliary grid is created with cells of size $r/\sqrt{2}$, where r is the minimum allowable distance between points. This ensures that there cannot be more than one point within a single cell or its neighbors. The process begins with the generation of the first random point within the permissible area, which is entered into the list of active points. For each active point, attempts are made to generate new

candidates: a point is randomly created at a distance of r to 2r from the initial point in a random direction, which geometrically corresponds to the selection in the ring area [4].

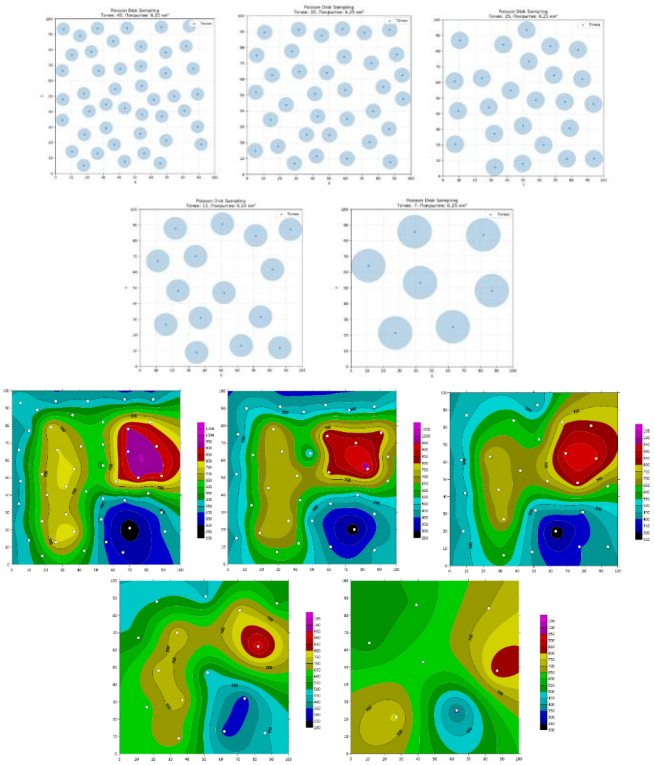


Figure 6. Spatial distribution of points obtained by Poisson sampling and further kriking interpolation with Poisson distribution of points (from 45 to 7 points). Interpolation based on these points showed a more stable distribution. Sharp boundaries and artifacts have less impact on the final map thanks to the uniform coverage of the study area.

For each such new point, it is checked whether the specified minimum distance to all other previously placed points is maintained, which is accelerated by referring not to all points, but only to neighboring grid cells. If the check is successful, the point is added to the final list and becomes a new active point; otherwise, the generation is repeated up to a specified number of iterations k. If, after k unsuccessful attempts, no new points can be added from the current active point, it is removed from the active list.

The algorithm continues until the required number of points is reached or the active list is empty. From a mathematical point of view, the distribution of points simulates weakly correlated positions with a Euclidean distance constraint, allowing for the approximation of uniform dense discretization, similar to the conditions of dense sampling in metric spaces [5].

This strategy is based on the Poisson sampling algorithm, which provides a more uniform and controlled distribution of points by introducing a minimum distance between them. For a given density and area, the number of points was scaled proportionally, taking into account the adjustment of the edge zones. The final distribution demonstrates a high degree of spatial uniformity (Fig. 6).

The use of Poisson disk sampling demonstrates significant advantages in terms of uniformity of coverage. Even at low densities (15 and 7 points), Poisson distribution minimizes clustering and provides a more balanced placement of points, which helps to reduce the number of artifacts and improve the reproduction of lithological zone boundaries. Interpolation maps constructed using the Kriging method are characterized by smoother transitions of Vs30 values and less pronounced local distortions compared to regular distribution.

RESULTS

3.1. ML interpolation of Vs30 model data with random distribution

Based on the results of the above studies and experiments, the best augmentation method and machine learning model for spatial interpolation of the Vs30 parameter in conditions of data scarcity were determined. For the purposes of comparative analysis and confirmation of the effectiveness of the selected augmentation and ML interpolation methods, a test experiment with random distribution was developed.

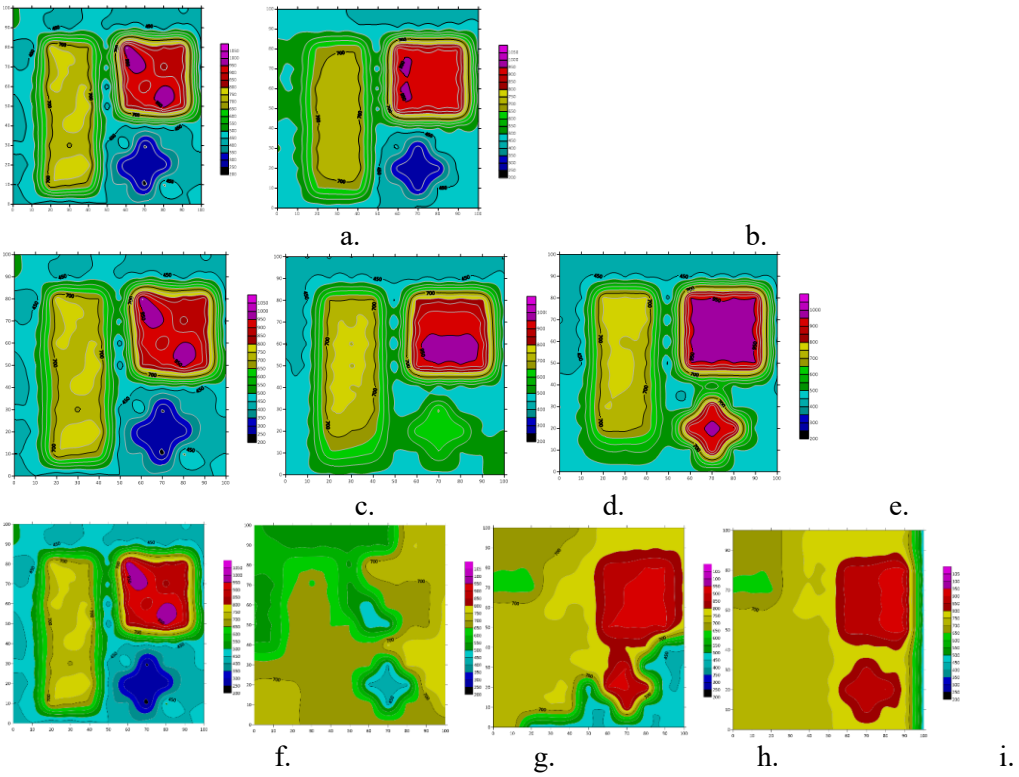


Figure 7. Vs30 interpolation under data scarcity for three training-set sizes. (a – reference Vs30 map, N = 25; b – XGBoost prediction, N = 25; c – reference map, N = 15; d – SVM prediction, N = 15; e – k-NN prediction, N = 15; f – reference map, N = 7; g – SVM prediction, N = 7; h – k-NN prediction, N = 7; i – XGBoost prediction, N = 7).

After generating 1,000 synthetic samples using the Residual (RF) Bootstrap method on random forest residuals and subsequently training ML models, we compared the resulting maps with reference maps for three training data volumes. Figure 7(a–b) shows the case of N = 25: panel a is the original (reference) Vs30 map, panel b is the XGBoost forecast. Visually, the model almost completely reproduces the structure of the reference: the contour lines, smoothness of gradients, and outlines of areas are practically identical; the added local details remain weak and do not introduce artifacts. Quantitatively, this is confirmed by 10-fold shuffle-split cross-validation (70/30): MSE = 575.10, RMSE = 43.87, MAE = 36.12, correlation $\rho = 0.97$, $R^2 = 0.94$. This means that with 25 initial points, RF-bootstrap provides XGBoost with a sufficient number of representative examples for a statistically sound and visually consistent reconstruction of the Vs30 field. For N = 15, the results are shown in Figure 7(c–e): c — reference map, d — SVR (RBF) forecast, e — k-NN forecast. Visually, SVR better retains global shapes and smooth gradients: the central “plateau” and peripheral depressions are outlined correctly, and the contours are smooth. In contrast, k-NN shows signs of overfitting: the ~950 m/s region is overly expanded, and the ~450 m/s regions acquire excessively sharp boundaries.

Cross-validation metrics confirm the advantage of SVR: CV-MSE = 2100, CV-MAE = 73, CV-RMSE = 123, ρ = 0.74, R^2 = 0.51 versus k-NN: CV-MSE = 2241, CV-MAE = 111, CV-RMSE = 226, ρ = 0.33, R^2 = −0.65. Consequently, with a moderately small set of measurements, SVR proves to be more stable and accurate than local neighbor methods.

Under conditions of extreme data scarcity (N = 7), the corresponding maps are shown in Figure X(f–i): f — reference, g — SVR, h — k-NN, i — XGBoost. Here, the differences between the models are particularly evident. SVR retains the lowest mean square error on cross-validation (CV-MSE = 672.47) and shows the best performance on external validation on 10 outlier points (MAE = 136.69, RMSE = 174.78, ρ = 0.77, while the low R^2 = 0.01 reflects the extremely limited explainable variance with such a small N and complex spatial structure). k-NN performs inconsistently (CV-MSE = 701.25, MAE = 196.83, RMSE = 265.02, ρ = 0.13, R^2 = −1.27), losing the ability to generalize correctly beyond the nearest neighbors. Despite its theoretical power, XGBoost in this scenario is prone to overfitting on synthetic augmentation: with a relatively low CV-MSE = 593.19, external metrics deteriorate (MAE = 211.85, RMSE = 267.38, ρ = 0.04, R^2 = −1.31), indicating a loss of generalization ability on unseen points. Overall, this shows that at N = 7, SVR provides the best combination of stability and accuracy, while k-NN and especially XGBoost are sensitive to the quality and structure of RF-bootstrap augmentation.

3.2. ML interpolation of Vs30 model data with Poisson distribution

In an experiment with 25, 15, and 7 initial points, expanded to 1000 using RF-bootstrap augmentation, the quality of the restoration of the spatial distribution of the target feature was evaluated using two machine learning models: gradient boosting on decision trees (XGBoost) and support vectors (SVR with RBF kernel).

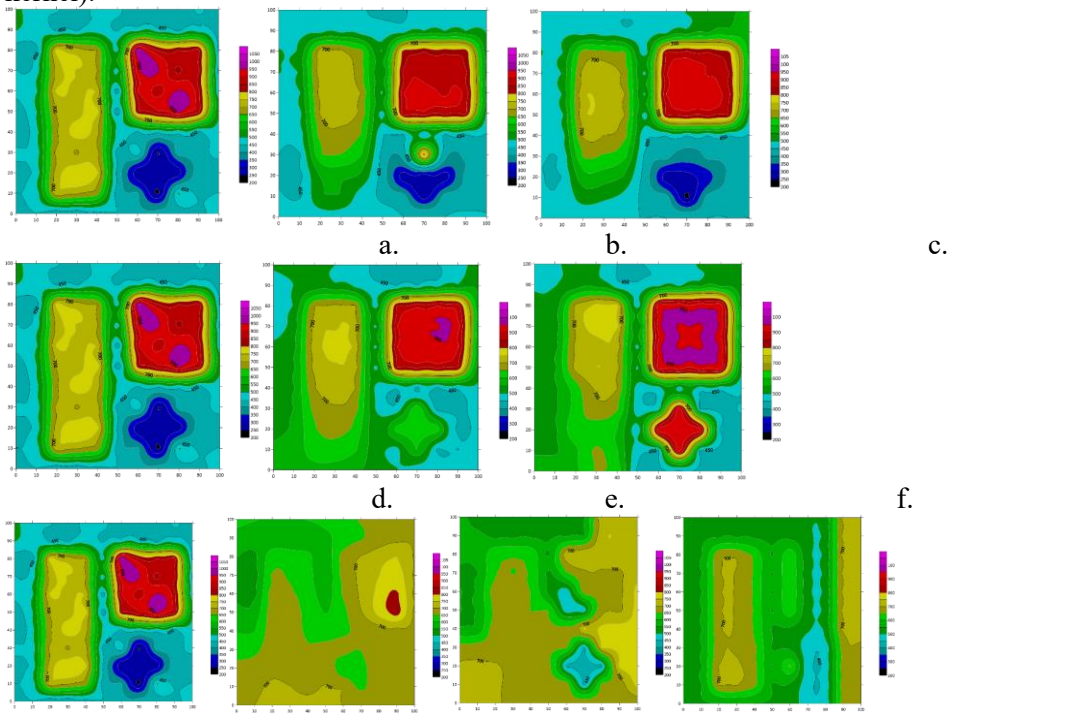


Figure 8. Interpolation Vs30 with three volumes of training data (a – original/reference map, N = 25; b – k-NN forecast, N = 25; c – SVM forecast, N = 25; d – original map, N = 15; e – SVR prediction, N = 15; f – k-NN prediction, N = 15; g – initial map, N = 7; h – SVR prediction, N = 7; i – k-NN prediction, N = 7; j – XGBoost prediction, N = 7).

After generating 1,000 synthetic samples using the Residual (RF) Bootstrap on random-forest residuals and subsequently training the ML models, we compared the resulting maps with the reference maps for three

training data volumes. Figure 8(a–c) shows the case of N = 25: panel a is the original (reference) Vs30 map, panel b is the k-NN prediction, and panel c is the SVR (RBF) prediction. Visually, SVR (c) better preserves the global shapes and smooth gradients than k-NN (b): the principal high-velocity domain and peripheral depressions are delineated with smoother contours and fewer local artifacts. Although k-NN aligns with the reference in broad terms, it exhibits staircase-like edges and mild over-expansion of high-velocity zones. Quantitatively (independent validation at N = 25), both SVR and XGBoost achieve a very high correlation ($\rho = 0.98$), but SVR attains lower errors (MAE = 42.44, RMSE = 47.88) and higher $R^2 = 0.93$ than XGBoost (MAE = 51.77, RMSE = 61.49, $R^2 = 0.88$), with CV-MSE = 1402.79 vs 1749.89 respectively—indicating a modest but consistent generalization edge for SVR at this data density.

For N = 15, the results are presented in Figure 8(d–f): d is the reference map, e the SVR prediction, and f the k-NN prediction. Visually, SVR (e) retains the global morphology and smooth transitions — central "plateau" and peripheral lows remain well-shaped—whereas k-NN (f) shows signs of overfitting. Cross-validation metrics corroborate this: SVR yields CV-MSE = 3080.12 and, on a 10-point external hold-out, MAE = 61.62, RMSE = 117.71, $\rho = 0.76$, $R^2 = 0.55$; k-NN has comparable CV-MSE = 3319.87 but degrades on the external set (MAE = 93.92, RMSE = 216.68, $\rho = 0.32$, $R^2 = -0.52$), evidencing weaker generalization under moderate data scarcity.

Under extreme data scarcity (N = 7), the corresponding maps are shown in Figure 8(g–j): g is the reference, h the SVR prediction, i the k-NN prediction, and j the XGBoost prediction. Here, model differences are most pronounced. SVR maintains the lowest cross-validated error (CV-MSE = 672.47) and the best external performance (MAE = 136.69, RMSE = 174.78, $\rho = 0.77$), while the near-zero $R^2 = 0.01$ reflects the inherently limited explainable variance with such a tiny N and complex spatial structure. k-NN behaves inconsistently (CV-MSE = 701.25, MAE = 196.83, RMSE = 265.02, $\rho = 0.13$, $R^2 = -1.27$), losing the ability to generalize beyond nearest neighbors. Despite its theoretical strength, XGBoost becomes vulnerable to overfitting on synthetic augmentation at N = 7: although CV-MSE = 5589.97 (lowest among the three) and MAE = 124.34 appear competitive, external metrics deteriorate (RMSE = 186.83, $\rho = 0.17$, $R^2 = -0.13$), indicating unstable predictions and loss of generalization on unseen points. Overall, SVR offers the best balance of stability and accuracy at N = 7, while k-NN, and especially XGBoost, are sensitive to the quality and structure of the RF-bootstrap augmentation.

DISCUSSION

When data is scarce, Poisson-disk sampling (PDS) provides more even coverage and fewer "gaps," which reduces Vs30 interpolation artifacts and improves metrics (MAE/RMSE/ρ/R^2) and visual map consistency (see Figs. 7–8). At N = 25, both regularized models agree with the reference ($\rho \approx 0.98$), but SVR consistently shows smaller errors and higher R^2 than XGBoost, indicating better generalization ability on augmented data. At N = 15, the advantage of SVR over k-NN increases: global shapes and smooth gradients are preserved, while k-NN is sensitive to the local configuration of points and "sharpens" the boundaries. In the extreme case of N = 7, the quality of all approaches declines; XGBoost shows signs of overfitting to synthetics (low external correlation and negative R^2), while SVR remains the least unstable, albeit with limited explained variance. In practice, this means: plan the network using PDS (as the default scheme), give priority to SVR for small N, and apply boosting models with caution and always check them on an external sample. The limitations of the study are related to the synthetic test bench and the nature of the augmentation; in further work, it is advisable to test real MASW/ReMi/SCPTu data and adaptively select r (and step) based on the local variogram and logistic constraints.

CONCLUSION

This paper considers an approach to planning the placement of observation points for interpolating the Vs30 parameter using the Poisson disk sampling (PDS) algorithm. The method has clear advantages over classical schemes—random and regular—especially when the number of observations is limited and there are areas of difficult access. PDS provides uniform coverage of the area at a given minimum interpoint distance,

which prevents clustering and increases the stability of interpolation estimates. A significant advantage of this approach is its adaptability: the introduction of a "radius of action" allows the planned point to be moved within the permissible area, bypassing protected or physically inaccessible areas without losing simulation quality.

Modeling on a 100×100 test geological map with four lithological zones showed that Poisson distribution of points provides a more accurate reconstruction of the Vs30 field even with small amounts of data (N = 15 and N = 7). This is confirmed by both geostatistical interpolation (kriging) and machine learning algorithms (Random Forest with RF-bootstrap, SVR, XGBoost, k-NN). In conditions of a lack of prior information, the PDS + SVR/XGBoost combinations demonstrated the best accuracy and stability metrics. Thus, PDS should be considered an effective tool for designing observation point layouts in seismic exploration, especially in densely built-up areas with limited access, where uniform coverage is critical. Its use improves the accuracy of Vs30 maps, optimizes fieldwork logistics, and can serve as the basis for automated engineering and geophysical survey planning systems.

REFERENCES

Allen, T.I., Wald, D.J. (2007). Topographic Slope as a Proxy for Seismic Site-Conditions (Vs30) and Amplification Around the Globe. *U.S. Geological Survey Open-File Report*.

Bridson, R (2007). Fast Poisson disk sampling in arbitrary dimensions. SIGGRAPH Technical Sketch.

Lagae, A., & Dutré, P (2008). A comparison of methods for generating Poisson disk distributions. *Computer Graphics Forum*, **27**(1), 114–129.

Matsuoka, M (2012). Amplification Capability and Geomorphologic Classification. *Tokyo Institute of Technology.*

Yusupov D. D., S. B. Khalbaev, J. Z. Kodirov, O. F. Zakirova, and T. U. Mamarozikov (2025). Vs30 interpolation under data scarcity: engineering–geological modeling and development prospects.